



## Appendix A

BsgI restriction endonuclease (RM fusion) coding sequence  
3108 nt (recognition sequence: GTGCAG)

```
ATGAATAGAG TAGAATCTAA AAAAAAATTA GAACAATTAG TTCAACAGTT CGAGAAGTAT
GAAAGTACAT ATAGCGCTTC GGATTATAAA GAGGCAACTT TAAGATCTAG TTTTITAGAT
CCCTTTTTTG AACTTTTTTG ATGGGAAATG CGCCCTGAAA GAATAACTAA TCCAGCAGAC
TTAGAAGTGA TTATAGAAGA AAGTTTAGAA ACGGAAAAAT CACTAAGTA TATAGATTAT
GTTTTTAAAA TTAATAGAAC GACTCAGTTT TTGGTAGAAG CTAAAAAGCC AGCTGAAAGT
CTTTCTAAAA AAGATCATAT TTTTCAGGCT AAAAGTTATG CATTACTAC GGAGATTCCA
TTTGTCATTT TAACAAATTT TAAAGAGTTC AGATTTTATG ACGTTTCAAC TGAACCTTTA
CACAATCAAC CGGATACAGA TAAAGTGGAA GAATATTGTT TTGATTATAA AGAATATGTT
CAAAACTTTG ATAAGTTATG GGAATTATTC AGCAGAGAAG CAGTTGCTAA CAGAAGTTTA
GCCAAGTTTT ATGCTAAAAG AAGAAATATA GTAGATAGTC CAGATTTAAT TTTTAACTT
AATTATCAAA TTGATAAAGG TGCATCATT CTGGATATAT CTTTCTTAAA AAATTTGAAA
ATATGGAGAA AATCATTAGC TGAAATATC TTTAATAATA ATTCACTTAA TGTTAACGTA
ATTAATGAAG TAGTTCAGAG AATATTAGAT AGACTGATAT TTATCCGTAT CATGAAGAC
AGAAATATTG AATCTAAAGA GTTTTTAAAA GAAATTGTAG AAATGCACGA ACAAGATAAT
TCGATTTTCA TGAAAAATGA ACTAGATAAA CTATGTATTG AATTAAATAA GAAATTTAAT
GGGTTAGTTT TCCATGACCA CACATTTGTT AACGAAGCGT TGATAGATAA CGAAATTTTA
ATAGTTATTA TTGACAATTT ATATTATCCA AAGTCTCCTT ATAACTTTAG ATTAATTAAG
CCAGAAATTT TAGGACGAAT ATTCGAGCAA TTTTGGGTG AAAAAATTGA AATAATCGAT
GGAAAAATAA CGTTAGGATT AAAAGATATT AATAAAAAAT CAGGAGGTGT TTATTATACA
CCTTCATATA TAGTTGAAAA AATAGTAGAA AATACATTAT CCAAAAAATT ACATAATGAT
ATTACTATTG AAAATTTAGA ACAGATAAAA ATAGCTGACA TAGCTTGTGG TTCAGGAAGC
TTTTTAATTT CATCATATAA ATATTTAATT GATAAATTTT AATATATTTA TTCCAAATGT
TCGGAAGCGG ATGTTCAAAC ATTAATTAGT AATAACTTAG TATTTATAGA CAATGGTAAA
TTAATGTAA CAATGGAACA TAAAAAGGGG ATACTTCAGC AAAATATTTT TGGGGTAGAT
ATAGATTCAC AAGCAATTCA GGTAGCGAAA TTAAGTCTTT ATATAACCAT GTTAGAAGAA
GGATACAGAG AAGGTACATT AAGACCTATA TTACCAGACT TAAATGATAA TATTAAACAT
GGTAACCTCA TAATAGATAA TGAAATTTTA TTTGAAGATG ATATAAATTA CGATATTGAT
GCAACATTAC CATTCGATTG GGAATATGCT TTTCTGATA TTATAGATAA CGGAGGTTTT
GATGTAATAT TAGGCAATCC ACCCTATATA AGAATTCAAA TTTTGAAGA GTTATATTGA
AAAGATGTAG TTAATTATTT GAAAAAATAA TACGTTTCTG CCGAAAAATT TAACTTTGAT
ATATATGTCG TGTTTATAGA AAAAGCATTG TCACTCTTGA ATGACCAGGG GATATTGGGA
TATATTGTGA TGAACAAATT TTTTACTACA CAATATGGAG AAAAATTGCG CGAGTTAATA
ACTTCACAAA AATTATTATA TGAAATCATT GATTTTGGAA TTAATGAAAT ATTTAATAAT
GCTACTACTT ATACTGTAT ATTAATTTTA GACAAACTA ATCCAGATGA AATAATTATT
GAAAGAGTGA TTGATTTAAA TACTTGGAAG GCTGGAGAAT CTTCAGATCG GAAAGTGGTA
GATCATACCG AATTCACTAG TACTCCTTGG TATTTATCAA GCAATACTGA TGAAGAAATT
TACAAATCTT TTGAAGAAAA TATGGTTTTA CTTGAAACCA TTAGTGATAG GGTTTTTGTT
GGTGTTTACA CAGACTGTGA TCCAGTATAT ATTTTAGAAG AAGTTTATGA AGAAGAAAA
TATTTATATT GTAAGTCAGA ATATACTACT GAAGTACACA AGTTTGAAAA AGATCATTTA
AAACCATTTT TAAAAGGTTT TCTAGATATA AAGAAATATA CTTTTTCAA TGTTAATAAG
TGTTTACTTT TCCCTTATAC CAATTCGGAA AATACTTCTG ATTTAATTCC CGAAACAAC
TACAAACAGT ATTTCCAGT AACATGGAAA TACTTAGAGT CTTGTAAAGA AAGATTAGCA
AAAAGAAAAA GTATTGAAAG AGAATTGGAT ATTAATCCGA ATTATAATGA GTGGTATAAA
TATATTTTACA AAAAGAATCA CACGAGGATG GACCAATTAA AAATAGTATT TCCTGCGATA
TCGAAGGGTA GTAGCTTTTG TTATGATTCG GATGGAGAGT ACTATTTTGT AGGAAGTGGT
GCTGGAGGCG GTGGTGGAGG CGCAATAGTC TTGCCAGATC AATCTGATTA TAATTATTTA
TCCTTACTTG GAATTCTAAA TTCAGAAGTA GTTTCATATC AAATGTGAAG AAGAGGTTCA
AAACATAAAG GTTCTTATTA TGGTGTAGAT AAAAAGAGAA TAGAAAATCT ATATGTGCCA
TTGATTAATG AGGATAATAA AAATTTATTT AGTAATATTT CAAAAATGGT AGCTCAAATT
CTTGATGCGT TTCAAAAAAT GCATCAAGCA GGGACAACGG ATGTTGGTAA AGAACAACCT
CAACAAGAA TAAAAATGCT TAATGCTAGA ATAAATGAGC TGGTATATAG ACTGTATAAT
TTACCAGTAG AATATAAAGA ATATATTAAA AATGCCTTAG AAAATTAA
```